CLAIMS

What is claimed is:

1. A method of enterprise web mining comprising the steps of:

collecting data from a plurality of data sources;

integrating the collected data;

generating a plurality of data mining models using the collected data; and

generating a prediction or recommendation in response to a received request for a recommendation or prediction.

2. The method of claim 1, wherein the collecting step comprises the steps of:

acquiring data from the plurality of data sources;

selecting data that is relevant to a desired output from among the acquired data;

pre-processing the selected data; and

building a plurality of database tables from the pre-processed selected data.

3. The method of claim 2, wherein the plurality of data sources comprises:

proprietary account or user-based data;

complementary external data;

web server data; and

web transaction data.

4. The method of claim 3, wherein the web server data comprises:

at least one of: web traffic data obtained by Transmission Control Protocol/Internet Protocol packet sniffing, web traffic data obtained from an application program interface of the web server, and a log file of the web server.

5. The method of claim 2, wherein the acquired data comprises a plurality of different types of data and integration step comprises the step of:

forming an integrated database comprising collected data in a coherent format.

6. The method of claim 5, wherein the model generating step comprises the steps of:

selecting an algorithm to be used to generate a model;

generating at least one model using the selected algorithm and data included in the integrated database; and

deploying the at least one model.

7. The method of claim 6, wherein the step of deploying the at least one model comprises the step of:

generating program code implementing the model.

8. The method of claim 7, wherein the step of generating an online prediction or recommendation comprises the steps of:

receiving a request for a prediction or recommendation;

scoring a model using data included in the integrated database;

generating a predication or recommendation based on the generated score; and

transmitting the predication or recommendation.

9. The method of claim 8, wherein the step of pre-processing the selected data comprises the step of:

performing, on the selected data, at least one of: data cleaning, visitor identification, session reconstruction, classification of web pages into navigation and content pages, path completion, and converting file names to page titles.

10. The method of claim 8, wherein the step of pre-processing the selected data comprises the step of:

collecting pre-defined items of data passed by a web server.

11. A computer program product for performing an enterprise web mining process in an electronic data processing system, comprising:

a computer readable medium;

computer program instructions, recorded on the computer readable medium, executable by a processor, for performing the steps of:

collecting data from a plurality of data sources;

integrating the collected data;

generating a plurality of data mining models using the collected data; and

generating a prediction or recommendation in response to a received request for a recommendation or prediction.

12. The computer program product of claim 11, wherein the collecting step comprises the steps of:

acquiring data from the plurality of data sources;

selecting data that is relevant to a desired output from among the acquired data;

pre-processing the selected data; and

building a plurality of database tables from the pre-processed selected data.

13. The computer program product of claim 12, wherein the plurality of data sources comprises:

proprietary account or user-based data;

complementary external data;

web server data; and

web transaction data.

14. The computer program product of claim 13, wherein the web server data comprises:

at least one of: web traffic data obtained by Transmission Control Protocol/Internet Protocol packet sniffing, web traffic data obtained from an application program interface of the web server, and a log file of the web server.

15. The computer program product of claim 12, wherein the acquired data comprises a plurality of different types of data and integration step comprises the step of:

forming an integrated database comprising collected data in a coherent format.

16. The computer program product of claim 15, wherein the model generating step comprises the steps of:

selecting an algorithm to be used to generate a model;

generating at least one model using the selected algorithm and data included in the integrated database; and

deploying the at least one model.

17. The computer program product of claim 16, wherein the step of deploying the at least one model comprises the step of:

generating program code implementing the model.

18. The computer program product of claim 17, wherein the step of generating an online prediction or recommendation comprises the steps of:

receiving a request for a prediction or recommendation;

scoring a model using data included in the integrated database;

generating a predication or recommendation based on the generated score; and

transmitting the predication or recommendation.

19. The computer program product of claim 18, wherein the step of pre-processing the selected data comprises the step of:

performing, on the selected data, at least one of: data cleaning, visitor identification, session reconstruction, classification of web pages into

navigation and content pages, path completion, and converting file names to page titles.

20. The computer program product of claim 18, wherein the step of pre-processing the selected data comprises the step of:

collecting pre-defined items of data passed by a web server.

21. A system for performing an enterprise web mining process, comprising:

a processor operable to execute computer program instructions; and

a memory operable to store computer program instructions executable by the processor, for performing the steps of:

collecting data from a plurality of data sources;

integrating the collected data;

generating a plurality of data mining models using the collected data; and

generating a prediction or recommendation in response to a received request for a recommendation or prediction.

22. The system of claim 21, wherein the collecting step comprises the steps of:

acquiring data from the plurality of data sources;

selecting data that is relevant to a desired output from among the acquired data;

pre-processing the selected data; and

building a plurality of database tables from the pre-processed selected data.

23. The system of claim 22, wherein the plurality of data sources comprises:

proprietary account or user-based data;

complementary external data;

web server data; and

web transaction data.

24. The system of claim 23, wherein the web server data comprises:

at least one of: web traffic data obtained by Transmission Control Protocol/Internet Protocol packet sniffing, web traffic data obtained from an application program interface of the web server, and a log file of the web server.

25. The system of claim 22, wherein the acquired data comprises a plurality of different types of data and integration step comprises the step of:

forming an integrated database comprising collected data in a coherent format.

26. The system of claim 25, wherein the model generating step comprises the steps of:

selecting an algorithm to be used to generate a model;

generating at least one model using the selected algorithm and data included in the integrated database; and

deploying the at least one model.

27. The system of claim 26, wherein the step of deploying the at least one model comprises the step of:

generating program code implementing the model.

28. The system of claim 27, wherein the step of generating an online prediction or recommendation comprises the steps of:

receiving a request for a prediction or recommendation;

scoring a model using data included in the integrated database;

generating a predication or recommendation based on the generated score; and

transmitting the predication or recommendation.

29. The system of claim 28, wherein the step of pre-processing the selected data comprises the step of:

performing, on the selected data, at least one of: data cleaning, visitor identification, session reconstruction, classification of web pages into navigation and content pages, path completion, and converting file names to page titles.

30. The system of claim 28, wherein the step of pre-processing the selected data comprises the step of:

collecting pre-defined items of data passed by a web server.

31. An enterprise web mining system comprising:

a database coupled to a plurality of data sources, the database operable to store data collected from the data sources;

a data mining engine coupled to the web server and the database, the data mining engine operable to generate a plurality of data mining models using the collected data;

a server coupled to a network, the server operable to:

receive a request for a prediction or recommendation over the network,

generate a prediction or recommendation using the data mining models, and

transmit the generated prediction or recommendation.

32. The system of claim 31, wherein the database comprises:

a plurality of database tables built from the collected data.

33. The system of claim 32, wherein the plurality of data sources comprises:

proprietary account or user-based data;

complementary external data;

web server data; and

web transaction data.

34. The system of claim 33, wherein the web server data comprises:

at least one of: web traffic data obtained by Transmission Control Protocol/Internet Protocol packet sniffing, web traffic data obtained from an application program interface of the web server, and a log file of the web server.

35. The system of claim 32, wherein the plurality of database tables forms an integrated database comprising collected data in a coherent format.

36. The system of claim 35, wherein the data mining engine is further operable to:

select an algorithm to be used to generate a model;

generate at least one model using the selected algorithm and data included in the integrated database; and

deploy the at least one model.

37. The system of claim 36, wherein the deployed model comprises program code implementing the model.

38. The system of claim 37, wherein the server is operable to generate a prediction or recommendation by scoring a model using data included in the integrated database and generating a predication or recommendation based on the generated score.

39. The system of claim 31, further comprising a data pre-processing engine pre-processing the selected data.

40. The system of claim 39, wherein the database comprises:

a plurality of database tables built from the pre-processed selected data.

41. The system of claim 40, wherein the plurality of data sources comprises:

proprietary account or user-based data;

complementary external data;

web server data; and

web transaction data.

42. The system of claim 41, wherein the web server data comprises:

at least one of: web traffic data obtained by Transmission Control Protocol/Internet Protocol packet sniffing, web traffic data obtained from an application program interface of the web server, and a log file of the web server.

43. The system of claim 40, wherein the plurality of database tables forms an integrated database comprising collected data in a coherent format.

44. The system of claim 43, wherein the data mining engine is further operable to:

select an algorithm to be used to generate a model;

generate at least one model using the selected algorithm and data included in the integrated database; and

deploy the at least one model.

45. The system of claim 44, wherein the deployed model comprises program code implementing the model.

46. The system of claim 45, wherein the server is operable to generate a prediction or recommendation by scoring a model using data included in the integrated database and generating a predication or recommendation based on the generated score.

47. The method of claim 46, wherein the data pre-processing engine pre-processes the selected data by performing, on the selected data, at least one of: data cleaning, visitor identification, session reconstruction, classification of web pages into navigation and content pages, path completion, and converting file names to page titles.

48. The method of claim 47, wherein the data pre-processing engine pre-processes the selected data by collecting pre-defined items of data passed by a web server.